P. ENT COOPERATION TREAT

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92bis.1 and Administrative Instructions, Section 422)

To:

GORDON, Richard, John, Albert
Barker Brettell
10-12 Priests Bridge
London SW15 5JE
ROYAUME-UNI

Date of mailing (day/month/year)
20 September 2000 (20.09.00)

Applicant's or agent's file reference
SP/4819 WO

IMPORTANT NOTIFICATION

International application No.
PCT/GB99/03737

International filing date (day/month/year)
09 November 1999 (09.11.99)

1. The following indications appeared on record concerning:

[ ] the applicant [ ] the inventor [X] the agent [ ] the common representative

Name and Address

PERKINS, Sarah
Stevens Hewlett & Perkins
Halton House
20/23 Holborn
London EC1N 2JD
United Kingdom

State of Nationality

State of Residence

Telephone No.
+44 20 7404 1955

Facsimile No.
+44 20 7404 1844

Teleprinter No.

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

[X] the person [X] the name [X] the address [ ] the nationality [ ] the residence

Name and Address

GORDON, Richard, John, Albert
Barker Brettell
10-12 Priests Bridge
London SW15 5JE
United Kingdom

State of Nationality

State of Residence

Telephone No.
+44 20 8392 2234

Facsimile No.
+44 20 8392 1858

Teleprinter No.

3. Further observations, if necessary:

4. A copy of this notification has been sent to:

[X] the receiving Office
[ ] the International Searching Authority
[X] the International Preliminary Examining Authority
[ ] the designated Offices concerned
[X] the elected Offices concerned
[ ] other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer
Marie-José Devillard

Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 20 July 2000 (20.07.00)	
International application No. PCT/GB99/03737	**Applicant's or agent's file reference** SP/4819 WO
International filing date (day/month/year) 09 November 1999 (09.11.99)	**Priority date** (day/month/year) 09 November 1998 (09.11.98)
Applicant GAMMERMAN, Alex et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

03 June 2000 (03.06.00)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland**	Authorized officer Olivia RANAIVOJAONA
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38

REPLACED BY ART 34 AMDT

PATENT COOPERATION TREATY

PCT

REC'D 29 DEC 2000
WIPO PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
RJG 1230		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/GB99/03737	09/11/1999	09/11/1998
International Patent Classification (IPC) or national classification and IPC G06K9/62		
Applicant ROYAL HOLLOWAY UNIVERSITY OF LONDON et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 5 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
03/06/2000	27.12.2000
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Lubach, E Telephone No. +49 89 2399 8991

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB99/03737

I. Basis of the report

1. This report has been drawn on the basis of *(substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments (Rules 70.16 and 70.17).)*:
Description, pages:

1-13 as originally filed

Claims, No.:

1-9 with telefax of 31/10/2000

Drawings, sheets:

1-3 as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

- ☐ the description, pages:
- ☐ the claims, Nos.:

☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	
	No:	Claims	1,5,6
Inventive step (IS)	Yes:	Claims	
	No:	Claims	2-4,7-9
Industrial applicability (IA)	Yes:	Claims	1-9
	No:	Claims	

2. Citations and explanations
see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

Ad VIII)

The iid assumption is defined on page 1 of the description as "the examples are generated from the same probability distribution independently of one another". It would appear to be rather tedious to try and separate classes that have the same probabilty distribution in feature parameter space. It would appear to make more sense that each from a particular class would be generated from the same probability distribution of that class (but that is not what the description says). But even if, for the sake of argument, this were assumed to be the case, the definition of the strangeness value in claim 1 is so broad and unspecific that the strangeness value interpreted from Backer and Duin is as good as the Applicants' "guess". The Applicants should seek to clarify the definition of strangeness value on the basis of the application as originally filed, i.e. without adding subject-matter.

Ad V)

The subject-matter of claim 1 reads onto conventional fuzzy classification typically conducted on computers (see "Statistische Patroonherkenning" by Backer and Duin, ISBN 90 6562 105 9, 1989, pages 129-137). Fuzzy labeling assigns membership values ("strangeness value") to an object for every potential classification and assigns a classification on the basis of the assigned labels. For each object, the sum of the membership values over all classes equals 1 (see p. 134). If an object **x** is allocated to a particular classification with membership **f(x)**, it follows in a trivial manner, that the misclassification fuzziness/likelihood is given by **1-f(x)** which is the sum of all remaining possible classifications including the second most likely one. Thus by way complement **f(x)** already serves as a strength of prediction estimation in which the second most likely classification is contained. Thus there appears nothing new in the subject matter of claim 1.

For the purpose of the discussion of novelty the iid restriction introduced by the Applicants does not impose any effective restrictions. To assume that the fuzzy set examples of the various classes in Backer and Duin are generated independently from respective class probability distributions, does not have any influence on the fuzzy labelling.

Similar comments apply to independent claims 5 and 6.

Dependent claim 2-4, 7 and 8 appear to concern commonplace variations of the subject-matter of the independent claims which are deemed to lie within the normal competence of those skilled in the art. These claims thus lack an inventive step.

The subject-matter of claim 9 is not rendered inventive by the presence of this method on a data carrier.

CLAIMS

1. Data classification apparatus comprising:

an input device for receiving a plurality of training classified examples and at least one unclassified example;

a memory for storing the classified and unclassified examples;

an output terminal for outputting a predicted classification for the at least one unclassified example; and

a processor for identifying the predicted classification of the at least one unclassified example

wherein the processor includes:

classification allocation means for allocating potential classifications to each unclassified example and for generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

assay means for determining a strangeness value for each classification set;

a comparative device for selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example, wherein the predicted classification output by the output terminal is the most likely allocated potential classification according to the strangeness values assigned by the assay means; and

a strength of prediction monitoring device for determining a confidence value for the predicted classification on the basis of the strangeness value of a classification set containing the second most likely allocated potential classification of the at least one unclassified example.

2. Data classification apparatus as claimed in claim 1, wherein the processor further includes an example valuation device which determines individual strangeness values for each training classified example and the at least one unclassified example having an allocated potential

classification.

3. Data classification apparatus as claimed in claim 2, wherein Lagrange multipliers are used to determine the individual strangeness values.

4. Data classification apparatus as claimed in claim 2, wherein the assay means determines a strangeness value for each classification set in dependence on the individual strangeness values of each example.

5. Data classification apparatus comprising:

an input device for receiving a plurality of training classified examples and at least one unclassified example;

a memory for storing the classified and unclassified examples;

stored programs including an example classification program;

an output terminal for outputting a predicted classification for the at least one unclassified example; and

a processor controlled by the stored programs for identifying the predicted classification of the at least one unclassified example

wherein the processor includes:

classification allocation means for allocating potential classifications to each unclassified example and for generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

assay means for determining a strangeness value for each classification set;

a comparative device for selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example, wherein the predicted classification output by the output terminal is the most likely allocated potential classification according to the strangeness values assigned by the assay means and

a strength of prediction monitoring device for determining a confidence value for the predicted classification on the basis of the strangeness value of the classification set containing the second most likely allocated potential classification of the at least one unclassified example.

6. A data classification method comprising:

inputting a plurality of training classified examples and at least one unclassified example;

identifying a predicted classification of the at least one unclassified example which includes,

allocating potential classifications to each unclassified example;

generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

determining a strangeness value for each classification set;

selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example wherein the predicted classification is the most likely allocated potential classification in dependence on the strangeness values;

determining a confidence value for the predicted classification on the basis of the strangeness value of the classification set containing the second most likely allocated potential classification for the at least one unclassified example; and

outputting the predicted classification for the at least one unclassified example and the confidence value for the predicted classification.

7. A data classification method as claimed in claim 6, further including determining individual strangeness values for each training classified example and the at least one unclassified example having an allocated potential classification.

8. A data classification method as claimed in any one of the preceding claims, wherein the selected classification set is selected without the application of any general rules determined from the training set.

9. A data carrier on which is stored a classification program for classifying data by performing the following steps:

generating a plurality of classification sets, each classification set containing a plurality of training classified examples and at least one unclassified example that has been allocated a potential classification;

determining a strangeness value for each classification set;

selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example wherein the predicted classification is the most likely allocated potential classification in dependence on the strangeness values; and

determining a confidence value for the predicted classification on the basis of the strangeness value of the classification set containing the second most likely allocated potential classification for the at least one unclassified example.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06K 9/62

A1

(11) International Publication Number: WO 00/28473

(43) International Publication Date: 18 May 2000 (18.05.00)

(21) International Application Number: PCT/GB99/03737

(22) International Filing Date: 9 November 1999 (09.11.99)

(30) Priority Data:
9824552.5 9 November 1998 (09.11.98) GB

(71) Applicant *(for all designated States except US)*: ROYAL HOLLOWAY UNIVERSITY OF LONDON [GB/GB]; Egham, Surrey TW20 0EX (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GAMMERMAN, Alex [RU/GB]; Royal Holloway University of London, Department of Computer Science, Egham, Surrey TW20 0EX (GB). VOVK, Volodya [UA/GB]; Royal Holloway University of London, Department of Computer Science, Egham, Surrey TW20 0EX (GB).

(74) Agents: PERKINS, Sarah et al.; Stevens Hewlett & Perkins, 1 Serjeants' Inn, Fleet Street, London, Greater London EC4Y 1NT (GB).

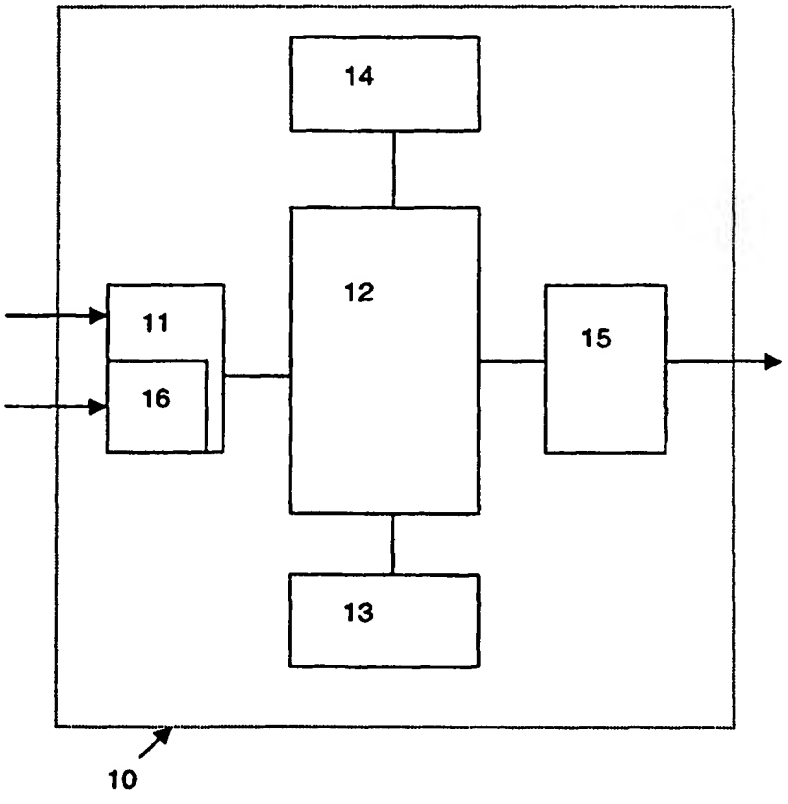
(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: DATA CLASSIFICATION APPARATUS AND METHOD THEREOF

(57) Abstract

The data classification apparatus and method is adapted to high–dimensional classification problems and provides a universal measure of confidence that is valid under the iid assumption. The method employs the assignment of strangeness values to classification sets constructed using classified training examples and an unclassified example. The strangeness values or p–values are compared to identify the classification set containing the most likely potential classification for the unclassified example. The measure of confidence is then computed on the basis of the strangeness value of the classification set containing the second most likely potential classification.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

Interr nal Application No
PCT/GB 99/03737

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06K9/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GAMMERMAN A ET AL: "Learning by transduction" UNCERTAINTY IN ARTIFICIAL INTELLIGENCE. PROCEEDINGS OF THE FOURTEENTH CONFERENCE (1998), PROCEEDINGS OF UNCERTAINTY IN ARTIFICIAL INTELLIGENCE (UAI-98), MADISON, WI, USA, 24-26 JULY 1998, pages 148-155, XP000869654 1998, San Francisco, CA, USA, Morgan Kaufmann Publishers, USA ISBN: 1-55860-555-X the whole document	1-9

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 February 2000	15/02/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Granger, B

1

PI

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION** see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
SP/4819 WO		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/GB 99/ 03737	09/11/1999	09/11/1998

Applicant

ROYAL HOLLOWAY UNIVERSITY OF LONDON et al.

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 2 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [ ] contained in the international application in written form.

   [ ] filed together with the international application in computer readable form.

   [ ] furnished subsequently to this Authority in written form.

   [ ] furnished subsequently to this Authority in computer readble form.

   [ ] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   [ ] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title**,

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1

   [X] as suggested by the applicant.

   [ ] None of the figures.

   [ ] because the applicant failed to suggest a figure.

   [ ] because this figure better characterizes the invention.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 99/03737

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06K9/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GAMMERMAN A ET AL: "Learning by transduction" UNCERTAINTY IN ARTIFICIAL INTELLIGENCE. PROCEEDINGS OF THE FOURTEENTH CONFERENCE (1998), PROCEEDINGS OF UNCERTAINTY IN ARTIFICIAL INTELLIGENCE (UAI-98), MADISON, WI, USA, 24-26 JULY 1998, pages 148-155, XP000869654 1998, San Francisco, CA, USA, Morgan Kaufmann Publishers, USA ISBN: 1-55860-555-X the whole document	1-9

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 February 2000	15/02/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Granger, B

1

Form PCT/ISA/210 (second sheet) (July 1992)

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06K 9/62 | A1

(11) International Publication Number: WO 00/28473

(43) International Publication Date: 18 May 2000 (18.05.00)

(21) International Application Number: PCT/GB99/03737

(22) International Filing Date: 9 November 1999 (09.11.99)

(30) Priority Data:
9824552.5 9 November 1998 (09.11.98) GB

(71) Applicant *(for all designated States except US)*: ROYAL HOLLOWAY UNIVERSITY OF LONDON [GB/GB]; Egham, Surrey TW20 0EX (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GAMMERMAN, Alex [RU/GB]; Royal Holloway University of London, Department of Computer Science, Egham, Surrey TW20 0EX (GB). VOVK, Volodya [UA/GB]; Royal Holloway University of London, Department of Computer Science, Egham, Surrey TW20 0EX (GB).

(74) Agents: PERKINS, Sarah et al.; Stevens Hewlett & Perkins, 1 Serjeants' Inn, Fleet Street, London, Greater London EC4Y 1NT (GB).

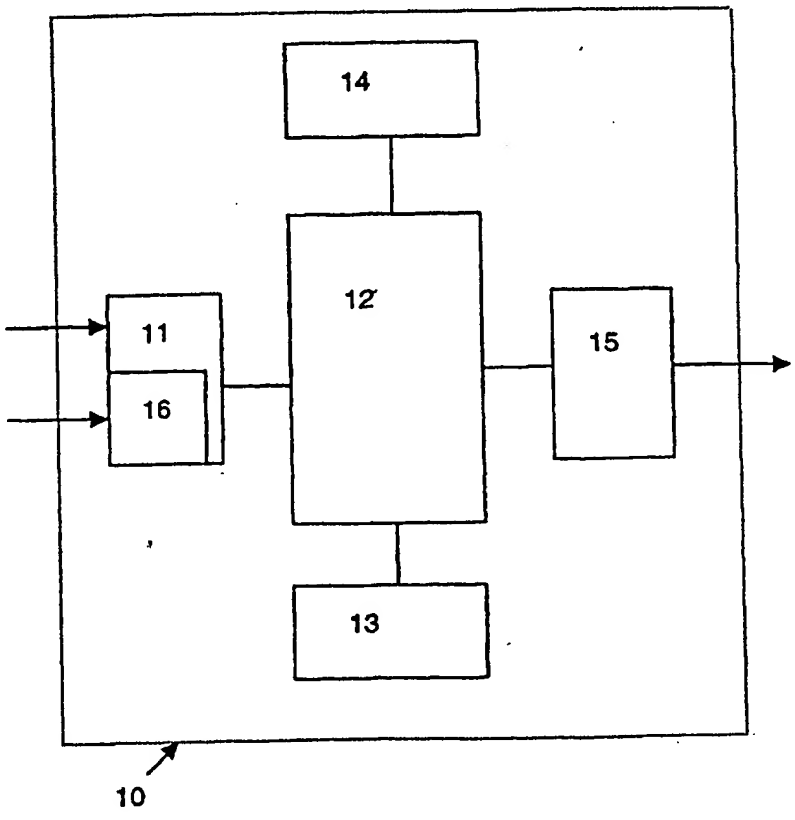
(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: DATA CLASSIFICATION APPARATUS AND METHOD THEREOF

(57) Abstract

The data classification apparatus and method is adapted to high–dimensional classification problems and provides a universal measure of confidence that is valid under the iid assumption. The method employs the assignment of strangeness values to classification sets constructed using classified training examples and an unclassified example. The strangeness values or p–values are compared to identify the classification set containing the most likely potential classification for the unclassified example. The measure of confidence is then computed on the basis of the strangeness value of the classification set containing the second most likely potential classification.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

DATA CLASSIFICATION APPARATUS AND METHOD THEREOF

BACKGROUND OF THE INVENTION

The present invention relates to data classification apparatus and an automated method of data classification thereof that provides a universal measure of confidence in the predicted classification for any unknown input. Especially, but not exclusively, the present invention is suitable for pattern recognition, e.g. optical character recognition.

In order to automate data classification such as pattern recognition the apparatus, usually in the form of a computer, must be capable of learning from known examples and extrapolating to predict a classification for new unknown examples. Various techniques have been developed over the years to enable computers to perform this function including, inter alia, discriminant analysis, neural networks, genetic algorithms and support vector machines. These techniques usually originate in two fields: machine learning and statistics.

Learning machines developed in the theory of machine learning often perform very well in a wide range of applications without requiring any parametric statistical assumptions about the source of data (unlike traditional statistical techniques); the only assumption made is the iid assumption (the examples are generated from the same probability distribution independently of each other). A new approach to machine learning is described in US5640492, where mathematical optimisation techniques are used for classifying new examples. The advantage of the learning machine described in US5640492 is that it can be used for solving extremely high-dimensional problems which are infeasible for the previously known learning machines.

A typical drawback of such techniques is that the techniques do not provide any measure of confidence in the predicted classification output by the apparatus. A typical user of such data classification apparatus just hopes that the accuracy of the results from previous analyses using benchmark datasets is representative of the results to be obtained from the

analysis of future datasets.

Other options for the user who wants to associate a measure of confidence with new unclassified examples include performing experiments on a validation set, using one of the known cross-validation procedures, and applying one of the theoretical results about the future performance of different learning machines given their past performance. None of these confidence estimation procedures though provides any practicable means for assessing the confidence of the predicted classification for an individual new example. Known confidence estimation procedures that address the problem of assessing the confidence of a predicted classification for an individual new example are ad hoc and do not admit interpretation in rigorous terms of mathematical probability theory.

Confidence estimation is a well-studied area of both parametric and non-parametric statistics. In some parts of statistics the goal is classification of future examples rather than of parameters of the model, which is relevant to the need addressed by this invention. In statistics, however, only confidence estimation procedures suitable for low-dimensional problems have been developed. Hence, to date mathematically rigorous confidence assessment has not been employed in high-dimensional data classification.

SUMMARY OF THE INVENTION

The present invention provides a new data classification apparatus and method that can cope with high-dimensional classification problems and that provides a universal measure of confidence, valid under the iid assumption, for each individual classification prediction made by the new data classification apparatus and method.

The present invention provides data classification apparatus comprising: an input device for receiving a plurality of training classified examples and at least one unclassified example; a memory for storing the classified and unclassified examples; an output terminal for outputting a predicted classification for the at least one unclassified example; and a processor for identifying the predicted classification of the at least one

unclassified example wherein the processor includes: classification allocation means for allocating potential classifications to each unclassified example and for generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification; assay means for determining a strangeness value for each classification set; and a comparative device for selecting a classification set containing the most likely allocated potential classification for at least one unclassified example, whereby the predicted classification output by the output terminal is the most likely allocated potential classification, according to the strangeness values assigned by the assay means.

In the preferred embodiment the processor further includes a strength of prediction monitoring device for determining a confidence value for the predicted classification on the basis of the strangeness value of a set containing the at least one unclassified example with the second most likely allocated potential classification.

With the present invention the conventional data classification technique of induction learning and then deduction for new unknown data vectors is supplanted by a new transduction technique that avoids the need to identify any all encompassing general rule. Thus, with the present invention no multidimensional hyperplane or boundary is identified. The training data vectors are used directly to provide a predicted classification for unknown data vectors. In other words, the training data vectors implicitly drive classification prediction for an unknown data vector.

It is important to note that with the present invention the measure of confidence is valid under the general iid assumption and the present invention is able to provide measures of confidence for even very high dimensional problems.

Furthermore, with the present invention more than one unknown data vector can be classified and a measure of confidence generated simultaneously.

In a further aspect the present invention provides data classification

apparatus comprising: an input device for receiving a plurality of training classified examples and at least one unclassified example; a memory for storing the classified and unclassified examples; stored programs including an example classification program; an output terminal for outputting a predicted classification for the at least one unclassified example; and a processor controlled by the stored programs for identifying the predicted classification of the at least one unclassified example wherein the processor includes: classification allocation means for allocating potential classifications to each unclassified example and for generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification; assay means for determining a strangeness value for each classification set; and a comparative device for selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example, whereby the predicted classification output by the output terminal is the most likely allocated potential classification, according to the strangeness values assigned by the assay means.

In a third aspect the present invention provides a data classification method comprising:

inputting a plurality of training classified examples and at least one unclassified example;

identifying a predicted classification of the at least one unclassified example which includes

allocating potential classifications to each unclassified example;

generating a plurality of classification sets each containing the plurality of training classified examples and the at least one unclassified example with an allocated potential classification;

determining a strangeness value for each classification set; and

selecting, according to the assigned strangeness values, a

classification set containing the most likely allocated potential classification; and outputting the predicted classification for the at least one unclassified example whereby the predicted classification output by an output terminal is the most likely allocated potential classification.

It will, of course, be appreciated that the above method and apparatus may be implemented in a data carrier on which is stored a classification program.

BRIEF DESCRIPTION OF THE DRAWINGS

An embodiment of the present invention will now be described by way of example only with reference to the accompanying drawings, in which:

Figure 1 is a schematic diagram of data classification apparatus in accordance with the present invention;

Figure 2 is a schematic diagram of the operation of data classification apparatus of Figure 1;

Figure 3 is a table showing a set of training examples and unclassified examples for use with a data classifier in accordance with the present invention; and

Figure 4 is a tabulation of experimental results where a data classifier in accordance with the present invention was used in character recognition.

DESCRIPTION OF PREFERRED EMBODIMENT

In Figure 1 a data classifier 10 is shown generally consisting of an input device 11, a processor 12, a memory 13, a ROM 14 containing a suite of programs accessible by the processor 12 and an output terminal 15. The input device 11 preferably includes a user interface 16 such as a keyboard or other conventional means for communicating with and inputting data to the processor 12 and the output terminal 15 may be in the form of a display monitor or other conventional means for displaying information to a user. The output terminal 15 preferably includes one or more output ports for connection to a printer or other network device. The data classifier 10 may be embodied in an Application Specific Integrated

Circuit (ASIC) with additional RAM chips. Ideally, the ASIC would contain a fast RISC CPU with an appropriate Floating Point Unit.

To assist in an understanding of the operation of the data classifier 10 in providing a prediction of a classification for unclassified (unknown) examples, the following is an explanation of the mathematical theory underlying its operation.

Two sets of examples (data vectors) are given: the training set consists of examples with their classifications (or *classes)* known and a test set consisting of unclassified examples. In Figure 3, a training set of five examples and two test examples are shown, where the unclassified examples are images of digits and the classification is either 1 or 7.

The notation for the size of the training set is l and, for simplicity, it is assumed that the test set of examples contains only one unclassified example. Let (X,A) be the measurable space of all possible unclassified examples (in the case of Figure 3, X might be the set of all 16 *x* 16 grey-scale images) and (Y,B) be the measurable space of classes (in the case of Figure 3, Y might be the 2-element set {1, 7}). Y is typically finite.

The confidence prediction procedure is a family $\{f_\beta:\beta\in(0, 1]\}$ of measurable mappings $f_\beta:(X \times Y)^l \times X \to B$ such that:

1. For any confidence level β (in data classification typically we are interested in β close to 1) and any probability distribution P in $X \times Y$, the probability that

$$y_{l+1} \in f_\beta(x_1, y_1, \ldots, x_l, y_l, x_{l+1})$$

is at least β, where $(x_1, y_1), \ldots, (x_l, y_l), (x_{l+1}, y_{l+1})$ are generated independently from P.

2. If $\beta_1 < \beta_2$, then, for all $(x_1, y_1, \ldots, x_l, y_l, x_{l+1}) \in (X \times Y)^l \times X$,

$$f_{\beta_1}(x_1, y_1, \ldots, x_l, y_l, x_{l+1}) \subseteq f_{\beta_2}(x_1, y_1, \ldots, x_l, y_l, x_{l+1})$$

The assertion implicit in the prediction $f_{\beta_1}(x_1, y_1, ..., x_l, y_l, x_{l+1})$ is that the true label y_{l+1} will belong to $f_{\beta_1}(x_1, y_1, ..., x_l, y_l, x_{l+1})$. Item 1 requires that the prediction given by f_β should be correct with probability at least β, and item 2 requires that the family $\{f_\beta\}$ should be consistent: if some label y for the $(l+1)$th example is allowed at confidence level β_1, it should also be allowed at any confidence level $\beta_2 > \beta_1$.

A typical mode of use of this definition is that some conventional value of β such as 95% or 99%, is chosen in advance, after which the function f_β is used for prediction. Ideally, the prediction region output by f_β will contain only one classification.

An important feature of the data classification apparatus is defining f_β in terms of solutions α_i, $i=1, \ldots, l+1$, to auxiliary optimisation problems of the kind outlined in US5640492, the contents of which is incorporated herein by reference. Specifically, we consider $|Y|$ completions of our data

$$(x_1, y_1), ..., (x_l, y_l), x_{l+1}$$

the completion y, $y \in Y$, is

$$(x_1, y_1), ..., (x_l, y_l), (x_{l+1}, y)$$

(so in all completions every example is classified).

With every completion

$$(x_1, y_1), ..., (x_l, y_l), (x_{l+1}, y_{l+1})$$

(for notational convenience we write y_{l+1} in place of y here) is associated the optimisation problem

$$\frac{1}{2}(w \cdot w) + C\left(\sum_{i=1}^{l+1} \xi_i\right) \longrightarrow \min \qquad (1)$$

(where C is a fixed positive constant)

subject to the constraints

$$y_i((x_i \cdot w) + b) \geq \xi_i, i = 1, ..., l+1 \qquad (2)$$

This problem involves non-negative variables $\xi_i \geq 0$, which are called *slack*

variables. If the constant C is chosen too large, the accuracy of solution can become unacceptably poor; C should be chosen as large as possible in the range in which the numerical accuracy of solution remains reasonable. (When the data is linearly separable, it is even possible to set C to infinity, but since it is rarely if ever possible to tell in advance that all completions will be linearly separable, C should be taken large but finite.)

The optimisation problem is transformed, via the introduction of Lagrange multipliers α_i, $i=1, \ldots, l+1$, to the dual problem: find α_i from

$$\sum_{i=1}^{l+1} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l+1} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \longrightarrow \max \qquad (3)$$

under the "box" constraints

$$0 \le \alpha_i \le C, \; i=1, 2, \ldots, l+1 \qquad (4)$$

The unclassified examples are represented, it is assumed, as the values taken by n numerical attributes and so $X = R^n$.

This quadratic optimisation problem is applied not to the attribute vectors x_i themselves, but to their images $V(x_i)$ under some predetermined function $V: X \rightarrow H$ taking values in a Hilbert space, which leads to replacing the dot product $x_i \cdot x_j$ in the optimisation problem (3)—(4) by the kernel function

$$K(x_i, x_j) = V(x_i) \cdot V(x_j)$$

The final optimisation problem is, therefore,

$$\sum_{i=1}^{l+1} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l+1} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \rightarrow \max$$

under the "box" constraints

$$0 \le \alpha_i \le C, \; i=1, 2, \ldots, l+1$$

this quadratic optimisation problem can be solved using standard packages.

The Lagrange multiplier α_i, $i \in \{1, \ldots, l+1\}$, reflects the "strangeness" of the example (x_i, y_i); we expect that α_{l+1} will be large in the wrong completions.

For $y \in Y$, define

$$d(y) := \frac{|\{i : \alpha_i \geq \alpha_{l+1}\}|}{l+1}$$

therefore $d(y)$ is the p-value associated with the completion y (y being an alternative notation for y_{l+1}). The confidence prediction function f, which is at the core of this invention, can be expressed as

$$f_\beta(x_1, y_1, \ldots, x_l, y_l, x_{l+1}) := \{y : d(y) > 1-\beta\}$$

The most interesting case is where the prediction set given by f_β is a singleton; therefore, the most important features of the confidence prediction procedure $\{f_\beta\}$ at the data $(x_1, y_1), \ldots, (x_l, y_l), x_{l+1}$ are:

- the largest $\beta=\beta_0$ for which $f_\beta((x_1, y_1), \ldots, (x_l, y_l), x_{l+1})$ is a singleton (assuming such a β exists);
- the classification $\mathbf{F}((x_1, y_1), \ldots, (x_l, y_l), x_{l+1})$ defined to be that $y \in Y$ for which $f_{\beta_0}((x_1, y_1, \ldots, (x_l, y_l), x_{l+1})$ is $\{y\}$.

$\mathbf{F}((x_1, y_1), \ldots, (x_l, y_l), x_{l+1})$ defined in this way is called the f-optimal prediction algorithm; the corresponding β_0 is called the confidence level associated with **F**.

Another important feature of the confidence estimation function $\{f_\beta\}$ at the data $(x_1, y_1), \ldots, (x_l, y_l), x_{l+1}$ is the largest $\beta=\beta_*$ for which $f_\beta((x_1, y_1), \ldots, (x_l, y_l), x_{l+1})$ is the empty set. We call $1-\beta_*$ the credibility of the data set $(x_1, y_1), \ldots, (x_l, y_l), x_{l+1}$; it is the p-value of a test for checking the iid assumption. Where the credibility is very small, either the training set $(x_1, y_1), \ldots, (x_l, y_l)$ or the new unclassified example x_{l+1} are untypical, which renders the prediction unreliable unless the confidence level is much closer to 1, than is $1-\beta_*$. In general, the sum of the confidence and credibility is between 1 and 2; the success of the prediction is measured by how close this sum is to 2.

With the data classifier of the present invention operated as

described above, the following menus or choices may be offered to a user:

1. Prediction and Confidence
2. Credibility
3. Details.

A typical response to the user's selection of choice 1 might be prediction: 4, confidence: 99%, which means that 4 will be the prediction output by the f-optimal **F** and 99% is the confidence level of this prediction. A typical response to choice 2 might be credibility: 100%, which gives the computed value of credibility. A typical response to choice 3 might be:

0	1	2	3	4	5	6	7	8	9
0.1%	1%	0.2%	0.4%	100%	1.1%	0.6%	0.2%	1%	1%

the complete set of p-values for all possible completions. The latter choice contains the information about $\mathbf{F}((x_1, y_1),\ldots,(x_l, y_l), x_{l+1})$ (the character corresponding to the largest p-value), the confidence level (one minus the second largest p-value) and the credibility (the largest p-value).

This mode of using the confidence prediction function f is not the only possible mode: in principle it can be combined with any prediction algorithm. If **G** is a prediction algorithm, with its prediction $y:=\mathbf{G}((x_1, y_1),\ldots,(x_l, y_l), x_{l+1})$ we can associate the following measure of confidence:

$$c(y) := \max\{\beta : f_\beta(x_1, y_1,\ldots, x_l, y_l, x_{l+1}) \subseteq \{y\}\}$$

The prediction algorithm **F** described above is the one that optimises this measure of confidence.

The table shown in Figure 4 contains the results of an experiment in character recognition using the data classifier of the present invention. The table shows the results for a test set of size 10, using a training set of size 20 (not shown). The kernel used was $K(x, y) = (x \cdot y)^3 / 256$.

It is contemplated that some modifications of the optimisation problem set out under equations (I) and (2) might have certain advantages, for example,

$$\frac{1}{2}(w \cdot w) + C\left(\sum_{i=1}^{l+1} \xi_i^2\right) \rightarrow \min,$$

subject to the constraints

$$y_i((x_i \cdot w) + b) = 1 - \xi_i, i = 1,...,l+1$$

It is further contemplated that the data classifier described above may be particularly useful for predicting the classification of more than one example simultaneously; the test statistic used for computing the p-values corresponding to different completions might be the sum of the ranks of αs corresponding to the new examples (as in the Wilcoxon rank-sum test).

In practice, as shown in Figure 2, a training dataset is input 20 to the data classifier. The training dataset consists of a plurality of data vectors each of which has an associated known classification allocated from a set of classifications. For example, in numerical character recognition, the set of classifications might be the numerical series 0—9. The set of classifications may separately be input 21 to the data classifier or may be stored in the ROM 14. In addition, some constructive representation of the measurable space of the data vectors may be input 22 to the data classifier or again may be stored in the ROM 14. For example, in the case of numerical character recognition the measurable space might consist of 16x16 pixellated grey-scale images. Where the measurable space is already stored in the ROM 14 of the data classifier, the interface 16 may include input means (not shown) to enable a user to input adjustments for the stored measurable space. For example, greater-definition of an image may be required in which case the pixellation of the measurable space could be increased.

One or more data vectors for which no classification is known are also input 23 into the data classifier. The training dataset and the unclassified data vectors along with any additional information input by the user are then fed from the input device 11 to the processor 12.

Firstly, each one of the one or more unclassified data vectors is provisionally individually allocated 24 a classification from the set of

classifications. An individual strangeness value α_i is then determined 25 for each of the data vectors in the training set and for each of the unclassified data vectors for which a provisional classification allocation has been made. A classification set is thus generated containing each of the data vectors in the training set and the one or more unclassified data vectors with their allocated provision classifications and the individual strangeness values α_i for each data vector. A plurality of such classification sets is then generated with the allocated provisional classifications of the unclassified data vectors being different for each classification set.

Computation of a single strangeness value, the p-value, for each classification set containing the complete set of training data vectors and unclassified vectors with their current allocated classification is then performed 26, on the basis of the individual strangeness values α_i determined in the previous step. This p-value and the associated set of classifications is transferred to the memory 13 for future comparison whilst each of the one or more unclassified data vectors is provisionally individually allocated with the same or a different classification. The steps of calculating individual strangeness values 25 and the determination of a p-value 26 are repeated in each iteration for the complete set of training data vectors and the unclassified data vectors, using different classification allocations for the unclassified data vectors each time. This results in a series of p-values being stored in the memory 13 each representing the strangeness of the complete set of data vectors with respect to unique classification allocations for the one or more unclassified data vectors.

The p-values stored in the memory are then compared 27 to identify the maximum p-value and the next largest p-value. Finally, the classification set of data vectors having the maximum p-value is supplied 28 to the output terminal 15. The data supplied to the output terminal may consist solely of the classification(s) allocated to the unclassified data vector(s), which now represents the predicted classification, from the classification set of data vectors having the maximum p-value.

Furthermore, a confidence value for the predicted classification is

generated 29. The confidence value is determined based on the subtraction of the next largest p-value from 1. Hence, if the next largest p-value is large, the confidence of the predicted classification is small and if the next largest p-value is small, the confidence value is large. Choice 1 referred to earlier, provides a user with predicted classifications for the one or more unknown data vectors and the confidence value.

Where an alternative prediction algorithm is to be used, the confidence value will be computed by subtracting from 1 the largest p-value for the sets of training data vectors and new vectors classified differently from the predicted (by the alternative method) classification.

Additional information in the form of the p-values for each of the sets of data vectors with respect to the individual allocated classifications may also be supplied (choice 3) or simply the p-value for the predicted classification (choice 2).

With the data classifier and method of data classification described above, a universal measure of the confidence in any predicted classification of one or more unknown data vectors is provided. Moreover, at no point is a general rule or multidimensional hyperplane extracted from the training set of data vectors. Instead, the data vectors are used directly to calculate the strangeness of a provisionally allocated classification(s) for one or more unknown data vectors.

While the data classification apparatus and method have been particularly shown and described with reference to the above preferred embodiment, it will be understood by those skilled in the art that various modifications in form and detail may be made therein without departing from the scope and spirit of the invention. Accordingly, modifications such as those suggested above, but not limited thereto, are to be considered within the scope of the invention.

CLAIMS

1. Data classification apparatus comprising:

an input device for receiving a plurality of training classified examples and at least one unclassified example;

a memory for storing the classified and unclassified examples;

an output terminal for outputting a predicted classification for the at least one unclassified example; and

a processor for identifying the predicted classification of the at least one unclassified example

wherein the processor includes:

classification allocation means for allocating potential classifications to each unclassified example and for generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

assay means for determining a strangeness value for each classification set;

a comparative device for selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example, wherein the predicted classification output by the output terminal is the most likely allocated potential classification according to the strangeness values assigned by the assay means; and

a strength of prediction monitoring device for determining a confidence value for the predicted classification on the basis of the strangeness value of a classification set containing the second most likely allocated potential classification of the at least one unclassified example.

2. Data classification apparatus as claimed in claim 1, wherein the processor further includes an example valuation device which determines individual strangeness values for each training classified example and the at least one unclassified example having an allocated potential

classification.

3. Data classification apparatus as claimed in claim 2, wherein Lagrange multipliers are used to determine the individual strangeness values.

4. Data classification apparatus as claimed in claim 2, wherein the assay means determines a strangeness value for each classification set in dependence on the individual strangeness values of each example.

5. Data classification apparatus comprising:

an input device for receiving a plurality of training classified examples and at least one unclassified example;

a memory for storing the classified and unclassified examples;

stored programs including an example classification program;

an output terminal for outputting a predicted classification for the at least one unclassified example; and

a processor controlled by the stored programs for identifying the predicted classification of the at least one unclassified example

wherein the processor includes:

classification allocation means for allocating potential classifications to each unclassified example and for generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

assay means for determining a strangeness value for each classification set;

a comparative device for selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example, wherein the predicted classification output by the output terminal is the most likely allocated potential classification according to the strangeness values assigned by the assay means and

a strength of prediction monitoring device for determining a confidence value for the predicted classification on the basis of the strangeness value of the classification set containing the second most likely allocated potential classification of the at least one unclassified example.

6. A data classification method comprising:

inputting a plurality of training classified examples and at least one unclassified example;

identifying a predicted classification of the at least one unclassified example which includes,

allocating potential classifications to each unclassified example;

generating a plurality of classification sets, each classification set containing the plurality of training classified examples and the at least one unclassified example with its allocated potential classification;

determining a strangeness value for each classification set;

selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example wherein the predicted classification is the most likely allocated potential classification in dependence on the strangeness values;

determining a confidence value for the predicted classification on the basis of the strangeness value of the classification set containing the second most likely allocated potential classification for the at least one unclassified example; and

outputting the predicted classification for the at least one unclassified example and the confidence value for the predicted classification.

7. A data classification method as claimed in claim 6, further including determining individual strangeness values for each training classified example and the at least one unclassified example having an allocated potential classification.

8. A data classification method as claimed in any one of the preceding claims, wherein the selected classification set is selected without the application of any general rules determined from the training set.

9. A data carrier on which is stored a classification program for classifying data by performing the following steps:

generating a plurality of classification sets, each classification set containing a plurality of training classified examples and at least one unclassified example that has been allocated a potential classification;

determining a strangeness value for each classification set;

selecting a classification set containing the most likely allocated potential classification for the at least one unclassified example wherein the predicted classification is the most likely allocated potential classification in dependence on the strangeness values; and

determining a confidence value for the predicted classification on the basis of the strangeness value of the classification set containing the second most likely allocated potential classification for the at least one unclassified example.

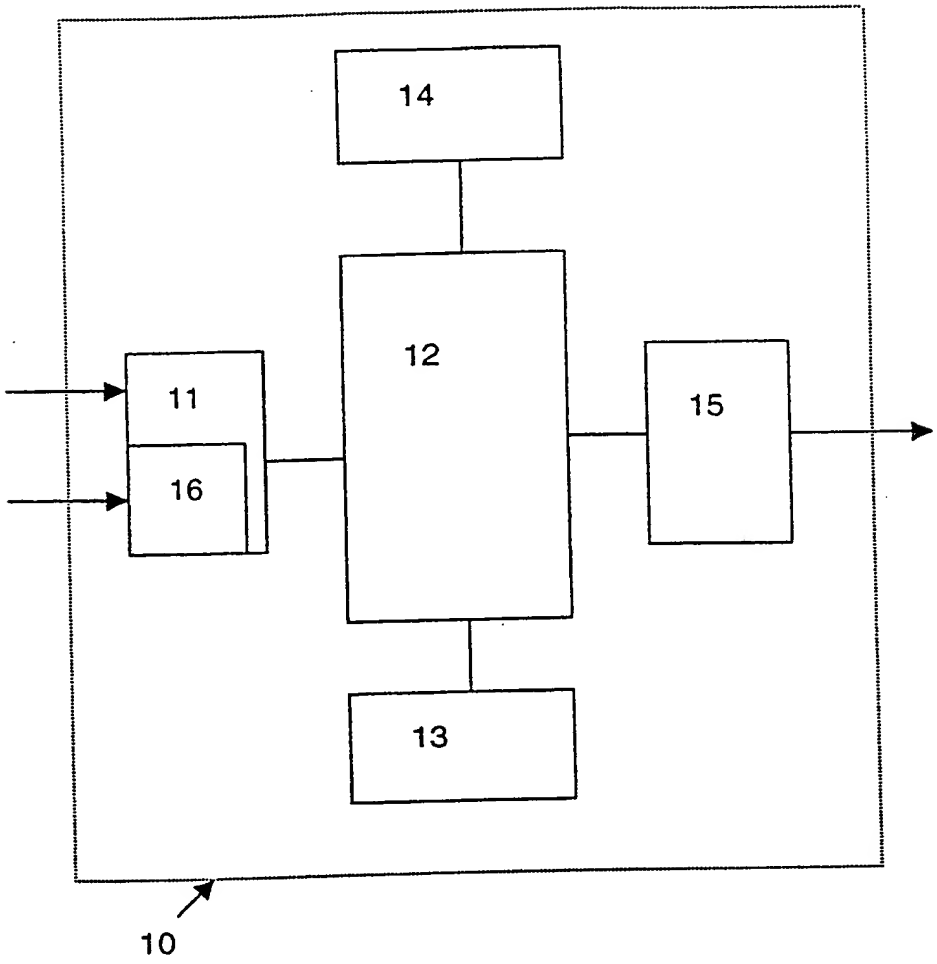


Figure 1

	Training Set					Test Set	
Example No.	1	2	3	4	5	1	2
Example	1	7	1	7	7	7	1
Classification	1	7	1	7	7	?	?

Figure 3

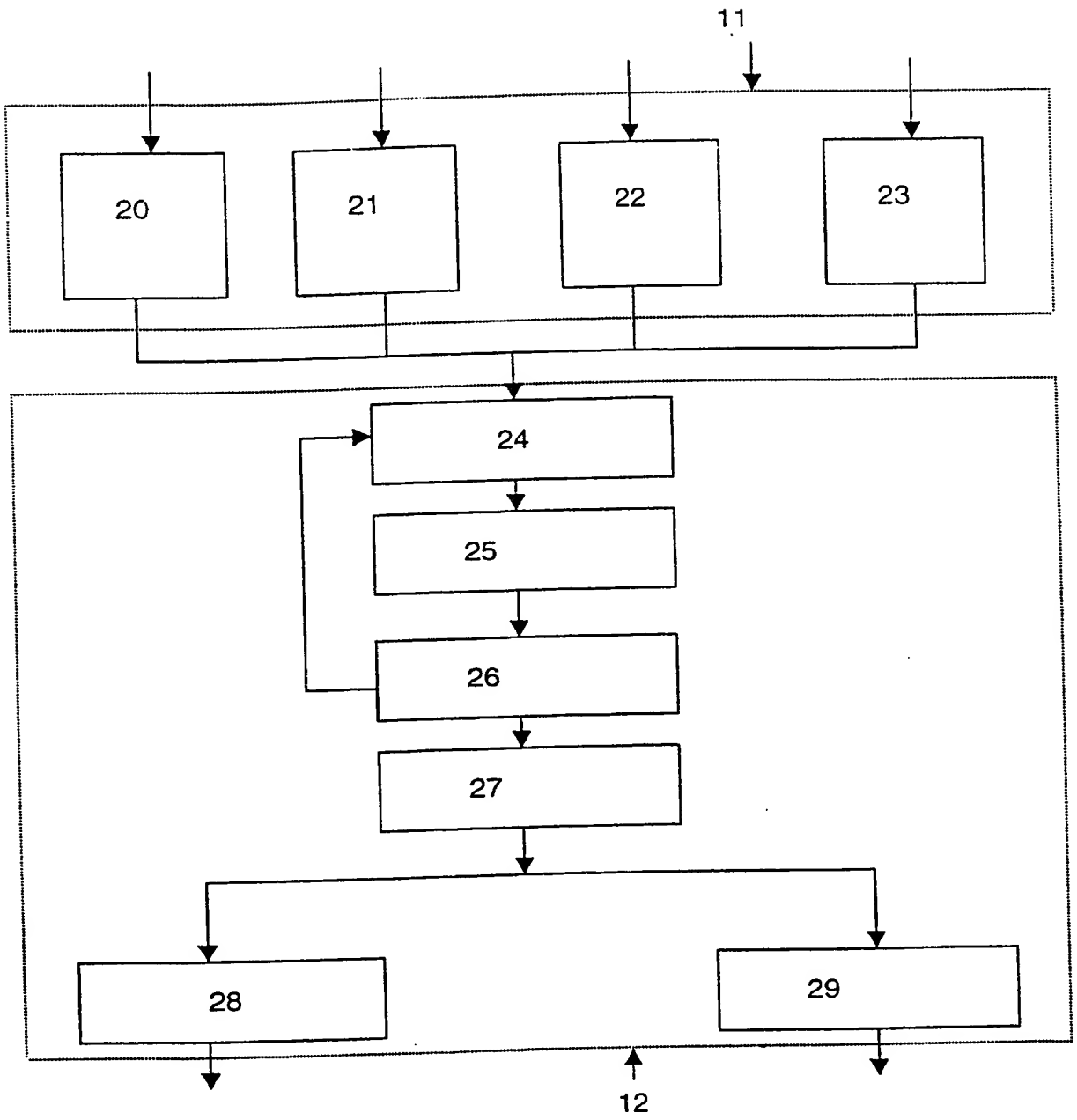


Figure 2

	Test Set									
Example No.	1	2	3	4	5	6	7	8	9	10
Example	1	7	7	1	7	1	1	1	7	7
True Class	1	7	7	1	7	1	1	1	7	7
Predicted Class	1	7	7	1	7	1	1	1	7	7
Confidence	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%
Credibility	19%	100%	100%	100%	100%	28%	100%	100%	100%	100%

Figure 4

INTERNATIONAL SEARCH REPORT

Interr nal Application No
PCT/GB 99/03737

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06K9/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GAMMERMAN A ET AL: "Learning by transduction" UNCERTAINTY IN ARTIFICIAL INTELLIGENCE. PROCEEDINGS OF THE FOURTEENTH CONFERENCE (1998), PROCEEDINGS OF UNCERTAINTY IN ARTIFICIAL INTELLIGENCE (UAI-98), MADISON, WI, USA, 24-26 JULY 1998, pages 148-155, XP000869654 1998, San Francisco, CA, USA, Morgan Kaufmann Publishers, USA ISBN: 1-55860-555-X the whole document	1-9

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 February 2000	15/02/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Granger, B

Form PCT/ISA/210 (second sheet) (July 1992)